REMARKS:

This paper is herewith filed in response to the Examiner's final Office Action mailed on March 3, 2009 for the above-captioned U.S. Patent Application. This office action is a rejection of claims 1-2, 5-14, 17-26, and 29-37 of the application.

More specifically, the Examiner has rejected claims 1-2, 5-12, 25-26, and 29-34 under 35 USC 101 as not falling within one of the four statutory categories of invention, and rejected claims 1-2, 5-14, 17-26, and 29-37 under 35 USC 103(a) as being unpatentable over Brecher (US 7,054,754) in view of Shanahan (US6,732,090). The Applicants disagree with the rejections.

Claims 1, 10-11, 13, 22-23, 25-26, 29, 31, and 33-36 have been amended for clarification. Support for the amendments can be found at least in paragraphs [0085], [0093], and [0098] of the published Application. No new matter is added.

Regarding the rejection under 35 USC 101 the Applicant notes that independent claims 1 and 25 have been amended to address the rejection, and the rejection is seen to be overcome. Further, the Applicants submit that, for at least the reason of their dependencies, the rejection of claims 2, 5-12, 26, and 29-34 under 35 USC 101 is also seen to be overcome. Thus, the rejection of all claims 1-2, 5-12, 25-26, and 29-34 under 35 USC 101 should be removed.

Regarding the rejection of claim 1 under 35 USC 103(a) the Applicants note that claim 1 has been amended to recite:

> A method to process a document, comprising: partitioning, with a tokenizer, document text separated by spaces into a plurality of tokens based on the spaces; identifying tokens to be ignored and not considered; determining that a first token considered of the plurality of tokens comprises a chemical name fragment, wherein determining comprises: examining syntax of the first token, examining context of the first token with respect to at least one adjacent token of the plurality of tokens, and taking into account the syntax and the context, applying to the first token a plurality of regular expressions, rules, and a plurality of dictionaries comprised of a prefix dictionary, and a suffix dictionary to recognize the chemical name fragment; **adding the recognized chemical name fragment to a vector of chemical name fragments, where**

> **the chemical name fragment is identified by a vector index variable**; combining the recognized chemical name fragment with at least one of the adjacent tokens that are determined to be a chemical name fragment into a complete chemical name, **where combining comprises: initializing the chemical name fragment vector index variable, incrementing the chemical name fragment vector index variable, where the incrementing continues at least until no chemical name fragments remain; setting a string combination to include the chemical name fragments identified by the initialized and incremented chemical name fragment vector index variables, and adding the string combination to a vector c as the complete chemical name**; assigning the complete chemical name with one part of speech; and storing in a memory the complete chemical name assigned with the one part of speech.

The Applicants note that claim 1 has been amended to recite, in part, features disclosed in at least paragraphs [0085], [0093], and [0098] of the published Application. More specifically, claim 1 as amended relates to adding a recognized chemical name fragment to a vector of chemical name fragments and using the vector for further processes recited in claim 1, including combining the chemical name fragments to form complete chemical names. The Applicants submit that neither Brecher nor Shanahan can be seen to disclose or suggest claim 1.

The Applicants note that for clarification of the arguments presented the Applicants have listed the following items pertaining to language recited in the claims and the rejections in the Office Action. These items are as follows:

item 1 – as in claim 1, **adding the recognized chemical name fragment to a vector of chemical name fragments, where the chemical name fragment is identified by a vector index variable**, and **setting a string combination to include the chemical name fragments identified by the initialized and incremented chemical name fragment vector index variables,**

item 2 – as in claim 1, **where combining comprises: initializing the chemical name fragment vector index variable, incrementing the chemical name fragment vector index variable, where the incrementing continues at least until no chemical name fragments remain; setting a string combination to include the chemical name fragments identified**

by the initialized and incremented chemical name fragment vector index variables, and adding the string combination to a vector c as the complete chemical name,
item 3 – as in claim 11, **wherein the plurality of dictionaries consists of the prefix dictionary, the suffix dictionary, and the negative dictionary**.

Reference items 1 and 2:

First, the Applicants note that exemplary embodiments of the invention relate to "a method and a computer program product for grouping [chemical] nomenclature into logical entities, **without the need to provide a large chemical dictionary or dictionaries**," (see the published application at par. [0029]). Whereas, Brecher relies on a seemingly large lexicon in which each known text string in Brecher is associated with terms in the lexicon (see Brecher col. 6, lines 35-45). The Applicants submit that Brecher fails to disclose or suggest any operation which can be seen to disclose or suggest at least where claim 1 relates to **adding the recognized chemical name fragment to a vector of chemical name fragments, where the chemical name fragment is identified by a vector index variable, and setting a string combination to include the chemical name fragments identified by the initialized and incremented chemical name fragment vector index variables**.

For example, with regards to combining nomTokens Brecher discloses:

> "The first sublist contains a nomToken of kTypeRoot ("phenacyl") that is preceded by another nomToken of kTypeRoot ("naphthoxy") that has exactly one entry in its attach-out list. Furthermore, the further preceding nomToken is of type kTypeUnknown and has a name ("p") that corresponds exactly to one of the entries in the locant map of "phenacyl". Therefore, the connection tables for "naphthoxy" and for "phenacyl" are combined, and a bond is indicated between the atom referenced in the attach-out map of "naphthoxy" and the atom referenced by the "p" entry in the locant map for "phenacyl"," (emphasis added), (col. 12, lines 11-21).

As stated above, Brecher discloses a method using lexicon entries for nomTokens of a type kTypeRoot in order to determine that "phenacyl" and "naphthoxy" are of the type kTypeRoot and as such they are somehow combinable. Thus, at least this process of combining in

Brecher is seen to relate to identifying the particular nomToken types as defined in the lexicon of Brecher. However, the Applicants can not find in all of Brecher any operation which can be seen to disclose or suggest adding recognized chemical name fragments to a vector of chemical name fragments, and using the vector to at least combine the recognized chemical name fragments into complete chemical names.

The Applicants contend that Brecher can not be seen to disclose or suggest at least where claim 1 recites in part:

> "**adding the recognized chemical name fragment to a vector of chemical name fragments, where the chemical name fragment is identified by a vector index variable**; combining the recognized chemical name fragment with at least one of the adjacent tokens that are determined to be a chemical name fragment into a complete chemical name, **where combining comprises: initializing the chemical name fragment vector index variable, incrementing the chemical name fragment vector index variable, where the incrementing continues at least until no chemical name fragments remain; setting a string combination to include the chemical name fragments identified by the initialized and incremented chemical name fragment vector index variables, and adding the string combination to a vector c as the complete chemical name**"

Further, although the Applicants do not agree that the proposed combination of Brecher and Shanahan is proper, the Applicants submit that Shanahan can not be seen to overcome at least the above stated shortfall of Brecher. The Applicants note that Shanahan discloses a use of aspect vectors to improve the precision of a query. In section F.3.2 Shanahan discloses:

> "The aspect vector is produced by analyzing a document's textual content using natural language processing in order to extract different facets of the document. In one embodiment, three facets of document content are examined (i.e., tokens (i.e., words), phrases, and rare words) to identify terms to retain. The retained terms are added to the recognized entity, in order to increase the precision of the query," (emphasis added), (col. 49, lines 51-57).

Further, Shanahan discloses:

> "At 4018, if the results are not accurate at 4016, then a determination is made whether a short run aspect vector has already been added to the query. If it has

> not already been added then a short run aspect vector using the document content and the entity as described above in section F.3.2 is generated at 4020. At, 4022 the aspect vector is added to the query and the node to which the query is pointing in the category organization is reset to the node that corresponds to its original categorization at 4024," (emphasis added), (col. 51, lines)

The Applicant submits that, as stated above, the aspect vector disclosed in Shanahan relates to improving the precision of a search query. The Applicants contend that in all of Shanahan there can not be found anything which can be seen to disclose or suggest where claim 1 recites in part:

> "**adding the recognized chemical name fragment to a vector of chemical name fragments, where the chemical name fragment is identified by a vector index variable**; combining the recognized chemical name fragment with at least one of the adjacent tokens that are determined to be a chemical name fragment into a complete chemical name, **where combining comprises: initializing the chemical name fragment vector index variable, incrementing the chemical name fragment vector index variable, where the incrementing continues at least until no chemical name fragments remain; setting a string combination to include the chemical name fragments identified by the initialized and incremented chemical name fragment vector index variables, and adding the string combination to a vector c as the complete chemical name**"

The Applicants contend that, for at least the reasons stated, even if the references were combined, which is not agreed to as proper, the proposed combination would still fail to disclose or suggest claim 1. Therefore, the rejection of claim 1 should be removed.

Further, the Applicants contend that for at least the reasons that independent claims 13, 25, and 35 recite features similar to claim 1, as stated above, the references cited can not be seen to disclose or suggest these claims. Therefore, the Examiner is respectfully requested to remove the rejections of all these claims 1, 13, 25, and 35.

Reference item 3:

In the Response to Arguments section of the Office Action dated December 5, 2008, the

Examiner states:

> "Applicants have amended the claims to include what was believed to be allowable. In particular, Applicants amended the claims to include the language "wherein a plurality of dictionaries consists of the prefix dictionary, the suffix dictionary and the negative dictionary. Applicant's explained previously that Brecher teaches the claimed dictionaries, however, Brecher uses more than just the claimed dictionaries and therefore, Brecher's invention does not consist of the prefix, suffix and negative dictionary. After further consideration, omission of the extra dictionaries would be obvious if the remaining dictionaries performed the same as before (In re Karlson, 136 USPO 184): "Omission of element and its function in combination is obvious expedient if remaining elements perform same functions as before." That is, since the extra dictionaries are used for some purpose outside the scope of the claim and the three disclosed dictionaries are used for the same purpose as claimed, then it would be obvious to omit the extra dictionaries plus their functions, since the 3 disclosed dictionaries would still be used for the same purpose as disclosed in the reference. Therefore, the claim limitation has been rejected in view of Brecher."

Although the Applicants do not agree with the Examiner's comments, the Applicants note that in this Response to Office Action the Applicants have removed this language, which was previously indicated as allowable in the prior Office Action dated August 6, 2008, from the independent claims 1, 20, 25, and 35. In addition, this language removed from independent claims 1, 20, 25, and 35 is now included in dependent claims 11, 23, 34, and 36 respectively.

Further, the Applicants submit that Brecher can not be seen to disclose or suggest at least where claim 11 recites, in part "where identifying tokens to be ignored comprises applying a negative dictionary to the plurality of tokens and **wherein the plurality of dictionaries consists of the prefix dictionary, the suffix dictionary, and the negative dictionary**."

As similarly stated above, the Applicants note that exemplary embodiments of the invention relate to "a method and a computer program product for grouping [chemical] nomenclature into logical entities, **without the need to provide a large chemical dictionary or dictionaries**," (see the published application at par. [0029]).

Brecher discloses:

> "In a specific embodiment, a fragment is determined to be meaningful ("recognized") if an exact match for the fragment is found in a dictionary of known text strings ("lexicon") that is maintained by the system," (emphasis added); and
>
> "Each known text string is associated in the lexicon with at least one data object known as a nomToken (FIG. 6). A nomToken includes the text of the known text string as its name and is described by Type and Subtype data members, which allow similar fragments to be grouped in accordance with two levels of similarity," (emphasis added), (col. 6, lines 35-45).

Here, Brecher discloses a dictionary or "lexicon" of known text strings, where each of the text strings is associated in the lexicon with a data object which Brecher refers to as a "nomToken." The Applicants submit that the lexicon of Brecher is seen to relate to a type of "large chemical dictionary," which, as similarly stated above, is explicitly taught away from in the pending application.

In addition, the Applicants submit that in Brecher it can be seen that there are numerous references to different nomToken types. It is noted that the Examiner appears to rely specifically on nomToken types kTypePrefix, kTypeSuffix, and kTypeUnknown to suggest the plurality of dictionaries which **consists** of the prefix dictionary, the suffix dictionary, and the negative dictionary, respectively, as in claim 11. However, the Applicants contend that, contrary to the Examiner's assertion, **more than these three nomTokens types are used for the purpose** of supporting the rejection of claim 11.

Related features as claim 11 depends from claim 1:

The Applicants note that claim 11 depends from claim 1. Thus, the relevant elements of claim 1, for which the plurality of dictionaries of claim 11 are used, are addressed accordingly as follows:

In regards to where claim 1 recites in part "identifying tokens to be ignored and not considered," the Examiner cites where Brecher discloses:

> "In this example, only one other environment is recognized in the sublists, and is found in the first sublist. The first sublist contains a nomToken of kTypeRoot ("phenacyl") that is preceded by another nomToken of kTypeRoot ("naphthoxy") that has exactly one entry in its attach-out list. Furthermore, the further preceding nomToken is of type kTypeUnknown and has a name ("p") that corresponds exactly to one of the entries in the locant map of "phenacyl". Therefore, the connection tables for "naphthoxy" and for "phenacyl" are combined, and a bond is indicated between the atom referenced in the attach-out map of "naphthoxy" and the atom referenced by the "p" entry in the locant map for "phenacyl". Accordingly, with respect to the three nomTokens involved, two are discarded and the resulting connection table is stored in the third, which leaves two nomTokens in the entire list (FIG. 7F), with exactly one nomToken in each sublist. As no other environments are recognized in either sublist, the two sublists are recombined," (emphasis added), (col. 12, lines 10-25).

The Applicants submit that, as cited, Brecher discloses the nomToken type **kTypeUnknown** "corresponds exactly to one of the entries in the locant map of [the **kTypeRoot** term] "phenacyl," [and] **Therefore** the connection tables for "naphthoxy" and for "phenacyl" are combined," (emphasis added). Further, Brecher, as cited, indicates that "**Accordingly, with respect to the three nomTokens involved, two are discarded,**" (emphasis added). Therefore, the Applicants submit that the identification of the kTypeRoot terms "phenacyl" and "naphthoxy" was integral to the method of Brecher where two of the nomTokens involved were discarded. The Applicants submit that, for at least these reasons, it can be seen that the listings in the lexicon of Brecher of the nomToken type's **kTypeUnknown and kTypeRoot** are applied in the rejection of claim 1 but not listed in the plurality of dictionaries of claim 11.

Further, in regards to where claim 1 recites in part "examining **syntax** of the first token," the Examiner cites where Brecher discloses:

> "When complete, the list of nomTokens is examined sequentially to determine whether any series of 2 . . . n adjacent nomToken names could be concatenated into a larger "buildable" nomToken (step 1040). This is due at least in part to the fact that a small number of chemical terms are commonly expressed with included punctuation, which the fragmentation process uses to divide the input name," and

> "For example, the phrase "mg/ml" could be interpreted as possibly unrecognized

nomTokens "mg" and "ml". Accordingly, "mgml" is recognized as a nomToken of type kTypeBuildable, which allows the two nomTokens "mg" and "ml" to be combined into one nomToken. The resulting nomToken of type kTypeBuildable is then converted to a nomToken of identical name and next-highest rank," (emphasis added), (col. 8, lines 29-42).

Thus, it can be seen that, as cited in Brecher, chemical terms expressed with punctuation and/or a forward slash (e.g., syntax) are recognized using a nomToken type kTypeBuildable. The Applicants submit that this recognized nomToken type of kTypeBuildable in the lexicon of Brecher is applied in the rejection of claim 1 but not listed in the plurality of dictionaries of claim 11.

Moreover, for where claim 1 recites in part "examining **context** of the first token with respect to at least one adjacent token of the plurality of tokens," the Examiner cites where Brecher discloses:

> "The list of nomTokens is examined for recognized environments. The first recognized environment is found when the list is examined for amino acids. No amino acids are found in the list, but one nomToken ("yl") of type kTypeEnderAminoAcid is present. Such a nomToken, being meaningful only in the context of amino acids, is not meaningful in this list that contains no amino acids. Accordingly, the nomToken of type kTypeEnderAminoAcid is converted to the next-highest-ranked nomToken of identical name, which in this case is of type kTypeSuffix and subtype kSubtypeYl," (emphasis added), (col. 11, lines 22-31).

In this example, Brecher discloses that the nomTokens are identified for recognized environments (e.g., context). Here, a term "yl" is identified as a type kTypeEnderAminoAcid. The Applicants submit that terms of this type are identified in the lexicon of Brecher. Further, the Applicants submit that it would be evident to a person of ordinary skill in the art that other terms corresponding to other nomToken types may be identified in different nomToken types listed in the lexicon. Thus, it can be seen that here, in Brecher, a nomToken type of **kTypeEnderAminoAcid** is applied in the rejection of claim 1 but not listed in the plurality of dictionaries of claim 11.

The Applicants contend that, for at least these reasons, Brecher clearly can not be seen to disclose or suggest at least where claim 11 now recites, in part "where identifying tokens to be ignored comprises applying a negative dictionary to the plurality of tokens and wherein the plurality of dictionaries **consists of the prefix dictionary, the suffix dictionary, and the negative dictionary**."

Therefore, for at least these reasons, the Applicants submit that claim 11 is patentably distinguishable from the references cited.

Further, for at least the reason that claims 23, 34, and 36 recite features similar to claim 11, as stated above, these claims are seen to be patentably distinguishable from the references cited.

Further, in the Interview Summary dated April 6, 2009 the Examiner states:

> "It is noted that Mr. Garrity pointed to claims 10 and 11 and explained that Brecher does not teach each of the listed character/character string. However, according to the present claim language, it recites where the characters comprise at least one of, and therefore each and every character/character string does not need to be taught."

The Applicants note that claim 10 as amended recites:

> "A method as in claim 8, **where the characters comprise upper case C, O, R, N and H**, and **where the characters comprise strings of lower case xy, ene, ine, yl, ane and oic**."

The Applicants submit that neither the characters of **upper case "C" nor "H"** is disclosed or suggested by Brecher. Further, Brecher as cited (originally in claim 11) is not seen to disclose or suggest **where the characters comprise strings of lower case xy, ene, ine, yl, ane and oic**, as in claim 10. The Applicants respectfully submit that for at least this reason the rejection of claim 10 should be removed.

Further, for at least the reasons that claims 22 and 33 recite features similar to claim 10, as

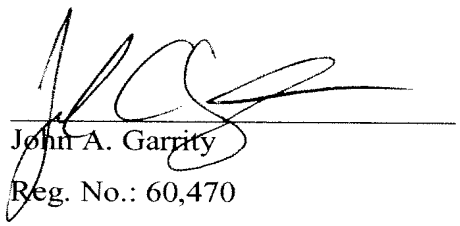
stated above, the Examiner is asked to remove the rejections of these claims.

In addition, for at least the reasons that claims 2 and 5-12, claims 14 and 17-24, claims 26 and 29-34, and claims 36-37 depend from claims 1, 13, 25, and 35, respectively, the rejections of these claims are seen to be improper and the rejections should be removed.

Further, the Applicants note that although not all the rejections are argued in this Response the Applicants do not acquiesce to these rejections.

It is respectfully submitted that all of the now-pending claims 1-2, 5-14, 17-26, 29-37 as presented herein are in condition for allowance. The Applicants respectfully request the rejections be withdrawn and these claims now be passed to issue. The Applicants' representative listed below welcomes the opportunity to resolve any matters that may remain via teleconference, at the Examiner's discretion.

Respectfully submitted:



6/3/2009
Date

John A. Garrity
Reg. No.: 60,470
Customer No.: 48237
HARRINGTON & SMITH, PC
4 Research Drive
Shelton, CT 06484-6212
Telephone: (203)925-9400
Facsimile: (203)944-0245
email:jgarrity@hspatent.com